

Mapping weather index-based insurance in Mali using spatial data analysis

Gerdien Meijerink^{a*}

Sjaak Conijn^b

Huib Hengsdijk^b

Karl Shutes^a

Marcel van Asseldonk^a

Michiel van Dijk^a

^a LEI, part of Wageningen UR, The Hague

^b PRI, part of Wageningen UR, Wageningen

* Corresponding author

Research Conference on Microinsurance, 2012, 11 April – 13 April, Enschede, The Netherlands.

Keywords: weather index based insurance, mapping, Mali, spatial analysis

1 Introduction

Weather is a significant factor for economic well-being in many countries of sub-Saharan Africa depending on agriculture as major livelihood, especially, in areas with predominantly rain fed agriculture (e.g. IBRD/WB, 2006). Under the threat of possible adverse weather conditions, poor farm households will often choose low-risk, low return activities, and avoid costly innovations that could increase productivity (e.g. Ruben et al., 2006). Financial institutions tend to restrict lending to farm households if adverse weather conditions might result in widespread defaults (Skees et al., 2006). The lack of access to credit restricts (poor) farm households access to agricultural inputs required for increasing productivity and overall rural development. Hence, different risk coping strategies at different scales exist that may be effective in reducing risk exposure in the short run, but hinder growth and rural development in the long-run.

As anticipated climate change is expected to increase yield variability, especially in Africa, the development of new formal risk management tools has been called for (Skees et al., 2006). One of the tools that received attention is the so-called ‘weather index-based insurance’ (WIBI) that distinguishes from traditional forms of insurance because compensation payments are based on an index that serves as proxy for losses rather than upon the individual losses of each policyholder. An index can be based on an objective measure such as rainfall in a certain period of the year that exhibits a strong correlation with the variable of interest, for example, crop yields (Hellmuth et al., 2009).

An important question is under which biophysical, climatic and socio-economic conditions WIBI is the most promising risk management option(?). An index-based insurance uses an

index that applies to a certain region, such as rainfall. If rainfall drops below a certain threshold, it is assumed that all farmers in this region are affected and will therefore receive a pay-out. However, crop yields are not affected only by rainfall. Of course, other biophysical factors play a role, such as soil type or the availability of irrigation water. These are usually also factored in when calculating the expected yield under normal circumstances (i.e. sufficient rain). Socioeconomic information is usually collected (typically through household surveys) after a target region has been identified based on agro-ecological and biophysical information. Mapping of socioeconomic information, however, could play a role much earlier, namely when identifying regions that could benefit from insurance.

There are several conditions that need to be fulfilled for WIBI to be successful. First, a given situation should face a risk of crop failure or low yields, which is not too large (otherwise the scheme would become too costly) or too small (otherwise farmers will not buy insurance). This risk can be expressed in terms of yield variability of a crop. In this paper we show how yield variability can be estimated using a combination of crop modelling and spatial data analysis. We use maize yields in Mali as an example.

Yield variability, however, not only depends on rainfall and agro-ecological factors, but also on socio-economic factors that influence management decisions of farmers. We use the concept of yield gaps as a proxy that reflects the management decisions of farmers, that are influenced by socio-economic conditions. Yield gap reflects the difference between potential yield (given agro-ecological conditions) and actual yield (the yield that farmers achieve in practice). Yield gaps are calculated the same way as yield variability, combining crop modelling and spatial data. Unfavourable socio-economic conditions are generally associated with large yield gaps, while under favourable socio-economic conditions yield gaps tend to be small. Assessment of the socio-economic indicators is helpful to identify which factors are the main bottleneck in closing the yield gap. We look at fertiliser use by farmers, market access and population density (which is used as a proxy for economic activity). We use spatial regression analysis to calculate the correlations between these factors.

This paper aims to show how spatial data analysis can help selecting appropriate areas for WIBI in a relatively cost-effective way. It allows a quick assessment of relatively large regions, avoiding the need for large field and household datasets. Needless to say, our suggested approach cannot fully substitute for field data. It is a first step, which needs to be followed up by a more hands-on approach in selected areas.

2 Material and methods

We apply two sets of methods: (i) a dynamic crop growth simulation model linked to spatial databases (ii) spatial regression models. The crop growth model is used to estimate potential maize yields as well as maize yield variability in Mali. Combining the spatially disaggregated data from the crop growth model with spatial datasets, we can calculate the yield gap. Spatial

regression models are used to calculate the correlation between yield gaps and socio-economic factors.

2.1 Explaining the yield gap

The yield gap is defined as the difference between the potential yield and the actual observed farmer's yield measured over a specified spatial and temporal scale of interest (Lobell et al., , 2009). It is mostly expressed in tonnes per hectare, but sometimes a fraction is also used. In this study, we define yield potential as 'the yield of a crop cultivar when grown in environments to which it is adapted with nutrients and water non-limiting and with pests, diseases, weeds, lodging, and other stresses effectively controlled (Evans and Fisher, 1999). Hence, it is a theoretical concept in which crop production is not restrained by any biophysical limitations other than a set of environmental factors that cannot be controlled through management, including solar radiation, temperature and plant characteristics. Van Ittersum and Rabbinge (1997) refer to these as growth-defining factors. They also distinguish two other sets of factors that can be controlled through management. Growth-limiting factors comprise water and nutrients, which are considered essential inputs for plant growth. If they are supplied in limited quantities, yields will decrease compared to the potential yield. Growth-reducing factors include pests, diseases, weeds, insects and pollutants. They will reduce crop growth and yield unless precautions are taken to prevent their impact (e.g. the use of pesticides, crop rotation and weed management). To achieve potential yields, optimal management of growth-limiting factors and growth-reducing factors is required, which is hard under field conditions.

The yield gap has two parts (Nin-Pratt et al., 2011). One part can never be closed because it represents the difference between a theoretical maximum (model simulation) or laboratory setting (research station and experimental fields) and the optimum that can be achieved in a non-perfect world. It is caused by random and uncontrollable environmental conditions (for example, extreme weather events, unanticipated seasonal conditions or unexpected pests that occur in reality but are not captured by the models), and the impact of specialised technologies and intensive practices that can be found only at test facilities.

The second part of the gap arises when farmers use practices and amounts of inputs that differ from what is needed to achieve the technical maximum farmer yield. In most cases, it is the direct reflection of a number of biophysical constraints (Table 1) that farmers face in practice, for example a limited availability of water and poor quality of planting material. Often also socio-economic factors can explain yield gaps, for example, a lack of knowledge of the production technology or economic constraints: the profit maximisation behaviour of farmers might lead to lower level of inputs than what would be used to reach the technical maximum yield. Moreover, market conditions and the diffusion of agricultural technology are determined by the interplay of a large number of socioeconomic factors that are mostly location-specific. Among others, these system-wide constraints include income, governance, market institutions, infrastructure and education. Conijn et al., (2011) provide an overview of these issues.

In this paper we will examine the link between yield gap and three socioeconomic indicators for which spatial data are available: market access, population density and fertiliser use. The importance of these factors has also been pointed out by the literature on development domains (Pender et al., 2006). Market access and infrastructure are critical determinants of regional comparative advantage. Areas with high market accessibility have better access to inputs such as fertiliser, pesticides and equipment as well as important services, mainly extension services and finance. Equally, market access and a high-quality road network will help farmers to link to value chains, facilitate exports, and reduce storage and transport costs. As all of this will contribute to higher yield, we expect a negative association between yield gap and market access.

The relation between population density and yield (and hence the yield gap) is not clear (Pender, 1999). Population pressure can increase the supply of labour that is available for agriculture. This reduces the costs of labour as opposed to the costs of land, and lead to more labour-intensive and less land-intensive agricultural production. Population pressure might also induce more capital-intensive production methods (e.g. the use of draft animals) and stimulate the adoption of more advanced technologies (e.g. improved seeds). Both would result in more production per hectare. On the other hand, population pressure might also lead to land degradation as a result of the cultivation of fragile lands, increased tillage and other forms of agricultural intensification, leading to lower yields. Pender (1999) found a negative relation between maize yield and population density in Honduras.

Figure 1. Comparison of yield measures

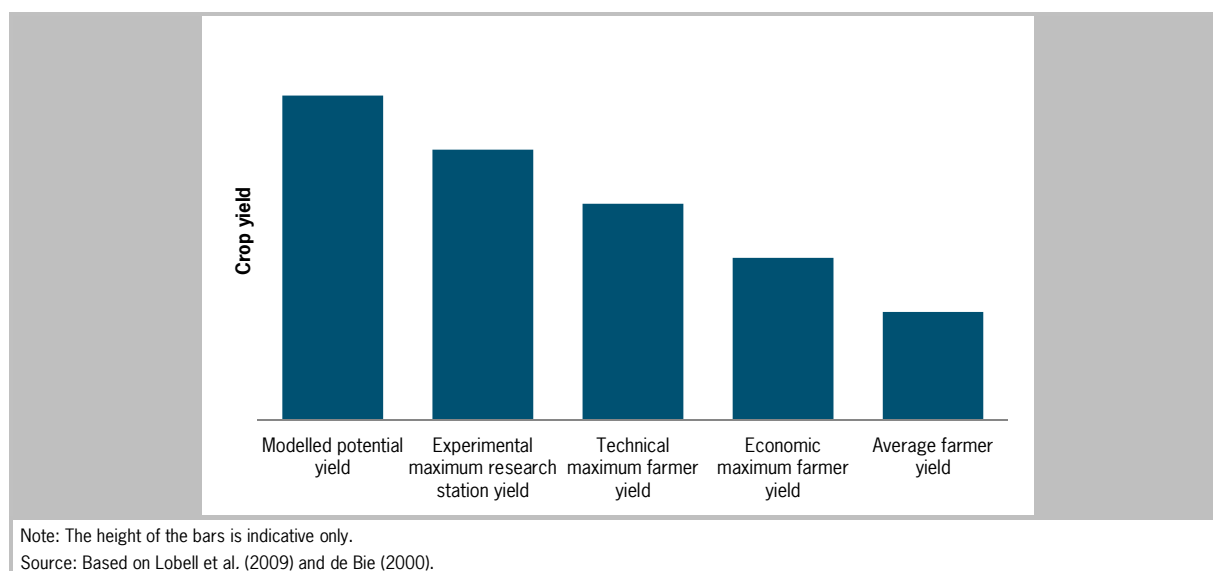


Table 1. Determinants of the yield gap

Biophysical constraints	Socioeconomic constraints
Insufficient water	Profit maximisation
Insufficient nutrients	Lack of knowledge of best practice management
Suboptimal planting (timing or density)	Risk avoidance strategies

Soil problems (e.g. salinity)	Inability to secure credit
Extreme weather events (e.g. floods, frost, hail)	Unpredictable prices of key inputs
Weed pressures	High transport costs
Pests and diseases	Distorted markets for fertiliser
Insect damage	Inefficiencies at harvest and storage problems
Source: based on Lobell et al. (2009).	

2.2 Yield gap estimate for Africa

Yield gaps for maize are determined per grid cell by combining estimated actual yield levels of around the year 2000 (1997-2003) with model-based estimates for yield potential. Actual yield levels are based on harvested maize areas and related maize yields provided by Monfreda and colleagues (2008) who combined statistical and remote sensing data. Yield potential for both irrigated and water-limited (rain-fed) maize areas were calculated using the crop model LINPAC (Appendix 1).

Maize yield and the yield gap – measured as the difference between yield potential¹ and actual yield and expressed in tonnes of dry matter (DM) per hectare – for the African continent are illustrated in Figure 2. In many grid cells, the yield gap is large; it varies from around 2.5 to over 12.5 tonnes per hectare per harvest. For some cells (about 2.5%), the model underestimates the maximum yield potential, resulting in a negative value for the yield gap. Table 2 provides descriptive statistics for the actual maize yield and yield gap as well as grid-level data on fertiliser use (Potter et al., 2010), population density (CIESIN, 2012) and market access (Verburg et al., 2011) that are used in the remainder of the paper.

Figure 2 (A) Average maize yield and (B) yield gap in tonnes of dry matter per hectare per harvest.

¹ There are two ways to calculate the yield gap: with or without taking into account the possibility of irrigation. We take into account the possibility of irrigation.

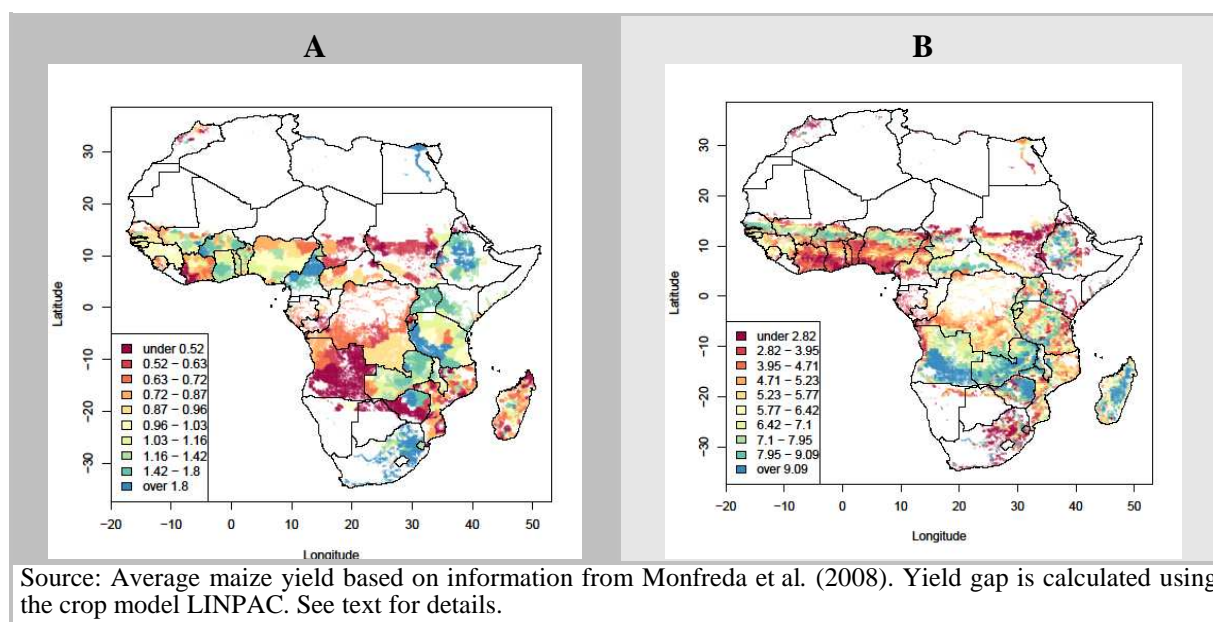


Table 2 Descriptive statistics

	Actual yield (tonnes DM per ha per harvest)	Yield gap (tonnes DM per ha per harvest)	Fertiliser (kg N per ha per year)	Market access	Population density people/km ²
Average	1.09	5.84	7.08	0.10	55
SD	0.75	2.56	26.45	0.19	163
Max	7.56	16.09	529.9	1.00	11,717
Min	0.02	-6.74	0	0	0

2.3 Yield variability estimate for Mali

The number of cropping cycles per year and the start(s) and end(s) of these cycles are calculated as function of soil moisture thresholds, i.e. adequate moisture conditions to start a cycle and prolonged drought conditions indicating the end, and the temperature sum required to complete crop development. If the length of the determined cropping cycle is too short relative to what the crop needs, a “no-cropping” situation is simulated which results in the simulation of only the soil water balance without a crop vegetation.

Soil water availability for the crop is determined by calculating infiltration, evapotranspiration and percolation. The soil profile is divided in two horizontal layers, i.e. from soil surface to the actual rooting depth and from actual rooting depth to a crop- or soil-specific maximum rooting depth. Water infiltrates in the soil as a result of precipitation plus (possible) irrigation minus runoff which is a function of soil texture, slope and precipitation/irrigation rate. Percolation equals the amount of water in excess of the maximum storage capacity of each soil layer and infiltrates into the next layer. Percolated water at the bottom layer is assumed to be lost for the crop. EvapoTranspiration (ET, mm) is calculated in two steps: (1) potential ET is calculated with the Penman-Monteith equation and divided over potential soil evaporation (E, mm) and crop transpiration (T, mm) and (2) actual ET is a function of this potential E and T and the soil water availability in the rooted soil layer. If the actual T falls below the

potential T, water stress occurs resulting in a proportional reduction of the crop biomass production and an acceleration of leaf senescence.

Additional water stress is modeled for maize around the period of flowering. The ratio of cumulative actual T over potential T shortly before and after flowering is computed and used to determine the sink strength of the grains to accumulate dry matter during grain filling. This ratio affects the dry matter production after flowering and is therefore positively correlated with ultimate grain yield.

The crop model operates with a daily time step to simulate the effects of day-to-day variability in weather conditions.

For each 5x5 arcminute grid cell in Mali and for each year in the period 1976 – 2006, the crop growth model is used to calculate daily soil water availability and crop growth as function of soil, weather and crop characteristics. These daily rates are integrated over time by the crop growth model resulting in, for example, crop dry matter yields per growth cycle and per year (= the sum of yields of all cycles in a year). To illustrate the risks of (extreme) low yields due to weather variability across Mali in the period 1976 - 2005, we map the percentage of years with yields lower than an (arbitrary set) yield threshold expressed as a percentage of the average yield.

2.4 Spatial data analysis

For our analysis we make use of publicly available spatial data. The development of high capacity and fast desk and laptop computers and the concomitant creation of geographic information systems has made it possible to explore georeferenced or mapped data as never before (Fischer and Getis, 2010). Coupled with open access policies such as those of the World Bank, there is an increasing availability of geo-coded data.

Spatial data is often obtained by remote sensing, which is the acquisition and analysis of data about an object or area acquired from a device that is not in contact with the object or area. Spatial statistics has increasingly become an integral part of the remote sensing process. The main issues facing researchers are that results differ in spatial scale and that typical study regions (landscapes) vary appreciably, even over short distances (Richards and Jia, 2006). We use different spatial regression models to analyse the yield gaps in Mali. In appendix 2 we discuss the spatial regression models used in this paper.

3 Mapping Rain and yield variability

Extreme weather events and natural disasters can trap rural households in poverty, impede development and drain a country's critical financial resources. Smallholders in developing countries are particularly vulnerable to such natural disasters.

The International Fund for Agricultural Development (IFAD) and the World Food Programme (WFP) have joined forces in the Weather Risk Management Facility (WRMF) to

improve the access of poor rural people to a range of financial services through the use of weather index-based insurance, a financial product based on local weather indices that are highly correlated with local crop yields. The WRMF focuses on four areas:

1. Building the capacity of local stakeholders for weather risk management by strengthening partnerships, offering technical assistance, and promoting knowledge exchange in the development and use of risk mitigation mechanisms, including weather index-based insurance (WII).
2. Improving weather services, infrastructure and data management for weather risk management, including the development of WII, national weather risk management, early warning systems and vulnerability analysis.
3. Supporting the development of an enabling environment by engaging with government partners and advocating national risk management frameworks and appropriate financial and weather risk-management strategies and policies.
4. Promoting inclusive financial systems for poor people in rural areas, including innovative delivery channels and client education, which lead to better planning for and coping with weather shocks.

Based on the analysis undertaken in the appraisal missions, IFAD and WFP have chosen Mali for implementation, taking a staged approach to the commencement of activities. In order to prepare the ground for implementation, additional research was needed. Wageningen UR was involved in this research in 2010 and 2011. Specifically, Wageningen UR contributed by assessing the feasibility for weather index-based insurance as a means of adaptation to climate change (Conijn et al., 2011; Meijerink and Shutes, 2011).

Figure 3 shows a map of Mali with the major towns and regions.

Figure 3 Map of Mali

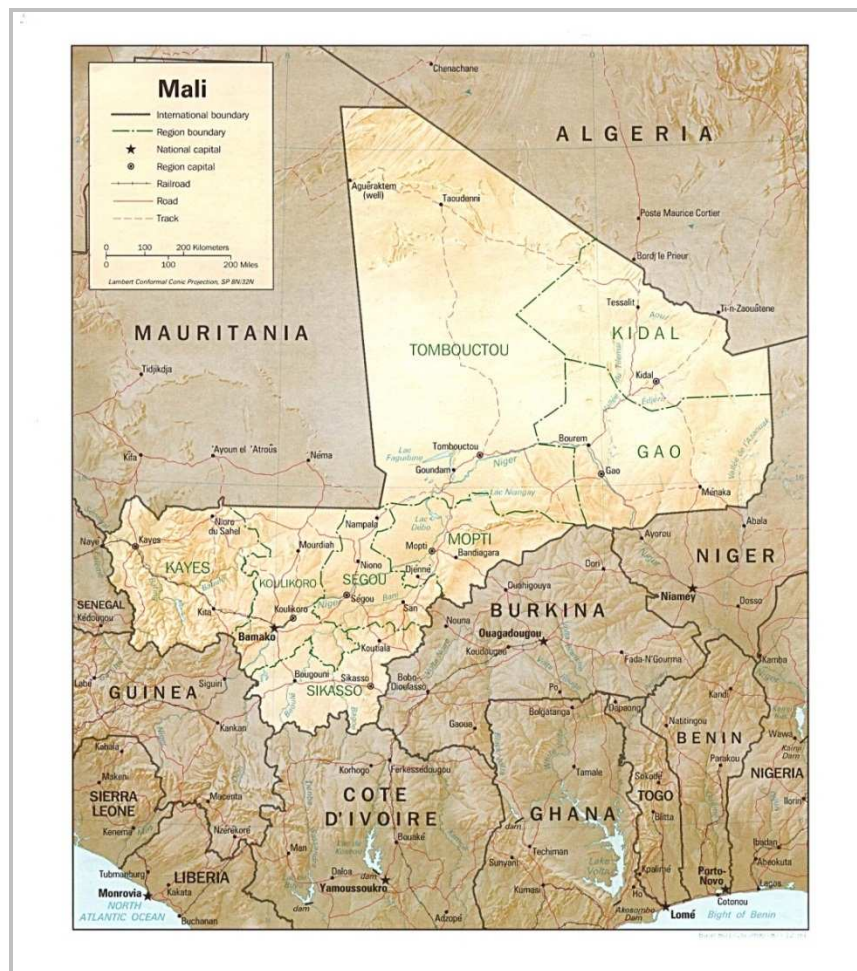


Figure 4 shows the total average rainfall per year across Mali during the period 1976 – 2005. Average rainfall shows a clear strong North-South gradient with less than 100 mm per year in the North and more than 1200 mm in the extreme South of Mali.

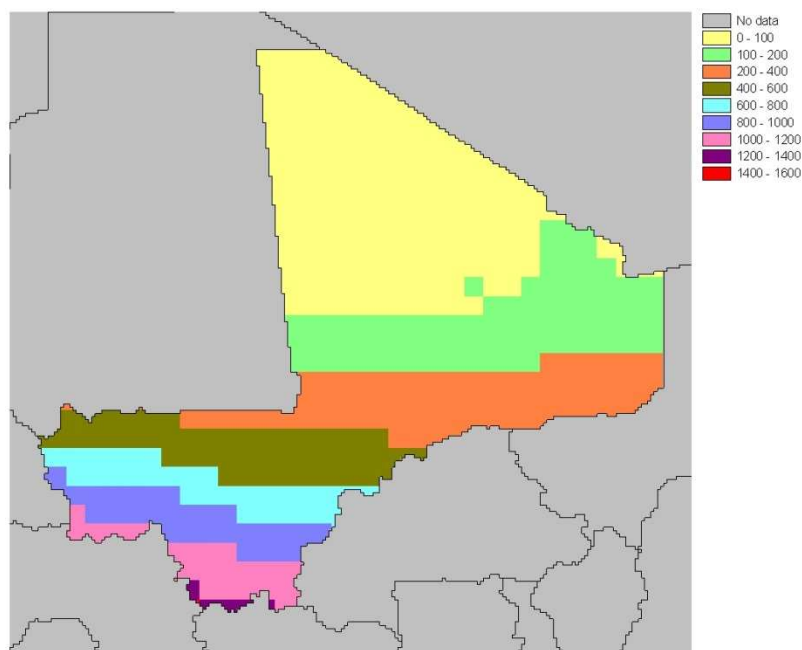


Figure 4. Total average rainfall in mm per year across Mali (period: 1976 – 2005).

The standard deviation runs from almost zero in the North to more than 150 mm in South and West Mali (Figure 5). Very high values (between 250 – 300 mm) are only found in one grid cell in the extreme South of Mali which also had the highest rainfall (> 1400 mm). The coefficient of variation of rainfall (standard deviation / average) ranges from almost 0% to over 100% during the period 1976 – 2005 (Figure 6). However, the coefficient of variation of rainfall in the majority of the crop land area in Mali is between 10 and 30%.

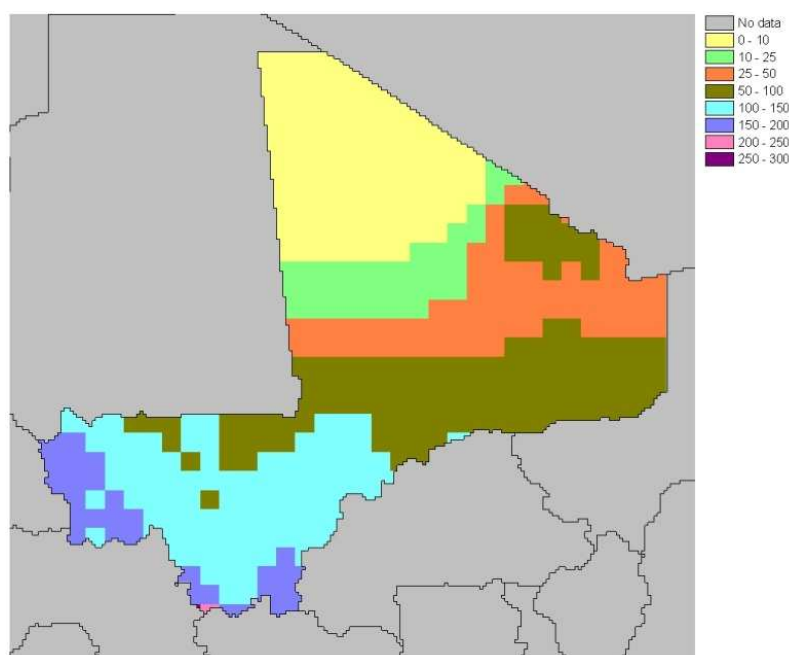


Figure 5. Standard deviation of total rainfall in mm per year in Mali (period: 1976 – 2005).

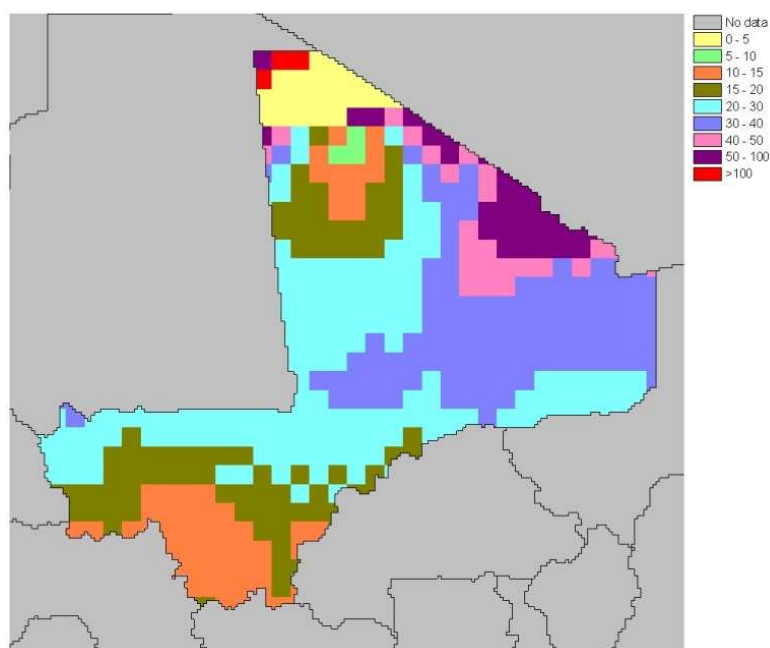


Figure 6. Coefficient of variation of total rainfall in % in Mali (period: 1976 – 2005).

Appendix 3 details how Figure 7 was derived.

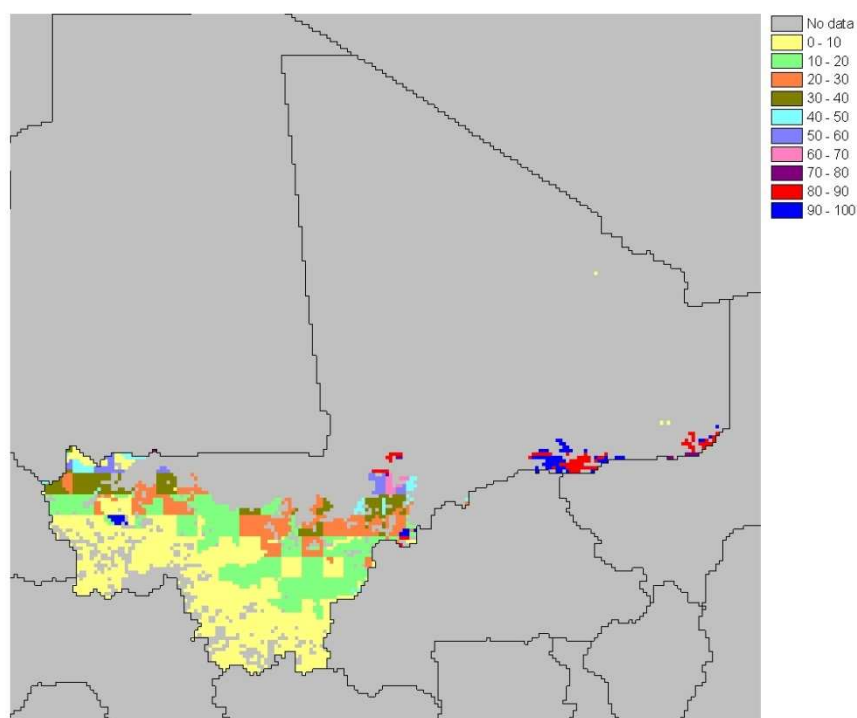


Figure 7. A map with percentage of years with low yield, here defined as < 40% of the 30-years average rain fed yield per grid cell used for growing maize in Mali around the year 2000 (compare with Figure 17; see Conijn et al. , 2011b).

In Figure 7 results are shown with a threshold of 40% using the same calculated yields (1976 – 2005; $n = 30$) for each grid cell and only for the maize cropping area (hence, it compares with Figure 17 in appendix 3). Logically, risks in Figure 7 are higher because the threshold is set at a higher value, which causes more years to comply with the condition of low yield ($< 40\%$ of the average).

Based on the 40% threshold (Figure 7), we analysed the maize area that might be suitable for weather index-based insurance within the period 1975-2005. The total harvested maize area in Mali is estimated at 160,000 ha around the year 2000 (Monfreda et al., 2008)². We assume that only those maize areas are of interest for a weather index-based insurance that face a probability of 10 to 30% that yields are 40% below the average yield in a grid cell, i.e. the combination of the probability classes 10-20 and 20-30 in Figure 7. This total suitable area is about 45,000 ha, but in most years the area with yields below 40% of the average is much less than 20,000 ha, and in some years the risk is even zero (Figure 8). This shows that unfavourable weather conditions causing low yields in a grid cell are spatially scattered in the selected maize region.

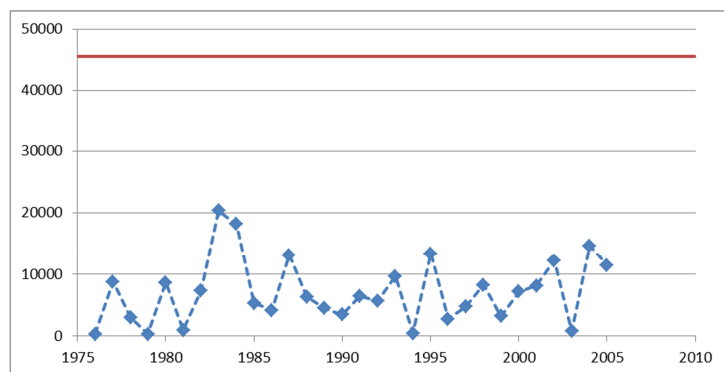


Figure 8. Maize area (ha) per year with a yield of 40% below the average grid cell yield. The red line indicates the total maximum maize area, here selected by probability classes 10-20 and 20-30 from Fig. 7.

4 Mapping socio-economic factors

We use the difference between potential yield and actual yield to identify areas that are characterised by their yield gaps. Socioeconomic factors (market access, fertiliser use and population density) are mapped to specify to what extent they can explain the yield gap. Such information can also be useful when targeting areas for insurance or upscaling pilot projects.

² National statistics from FAOSTAT report for the same period a harvested area of maize of circa 275,000 ha, increasing towards circa 400,000 ha in recent years. The difference between FAOSTAT and Monfreda et al (2008) is due to the use of different statistical databases.

Figure 9. Map with percentage of years with low yield, here defined as <40% of the 30-years average rain-fed yield per grid cell used for growing maize in southern Mali around the year 2000

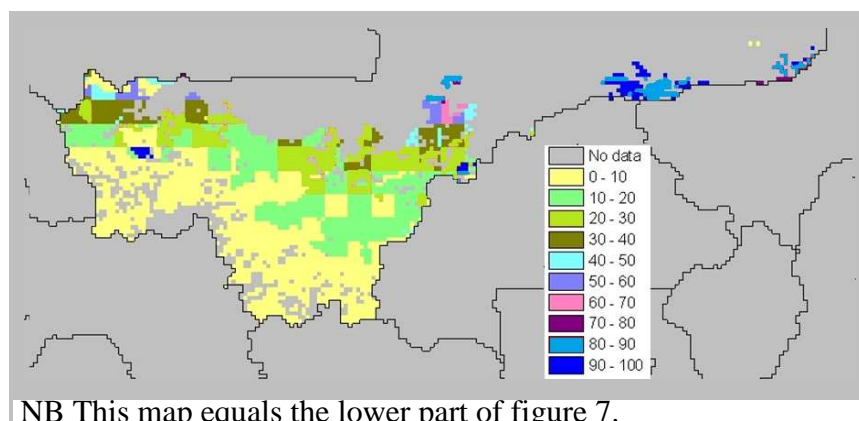


Figure 9 can be seen as a spatial risk distribution for maize production. Based on the 40% threshold, this maize area might be suitable for weather index-based insurance: for instance, only those maize areas that face a 10–30% probability that yields will be 40% below the average yield in a grid cell, that is, the combination of the probability classes 10-20 and 20-30 in Figure 9 (both light green). Only southern Mali is shown, because maize is hardly grown in the rest of Mali.

This can be combined with other spatial information, as described elsewhere in this report. One important question in targeting weather index-based insurance is whether yield is influenced by other factors than climate alone. It makes a big difference whether farmers apply fertiliser, for instance. Access to markets may also be important: to buy inputs farmers need access to markets, and if farmers can easily sell produce, they may invest more in them. Population density may be also important, as it determines labour availability. We will explore these factors with respect to Mali and focus on the light green areas identified in Figure 9.

We use a spatial econometric approach (see appendix 2 for details) to show how areas are clustered and how different factors (yield gap, fertiliser, market access and population density) are related.

Figure 10. Yield gaps spatially correlated for southern Mali

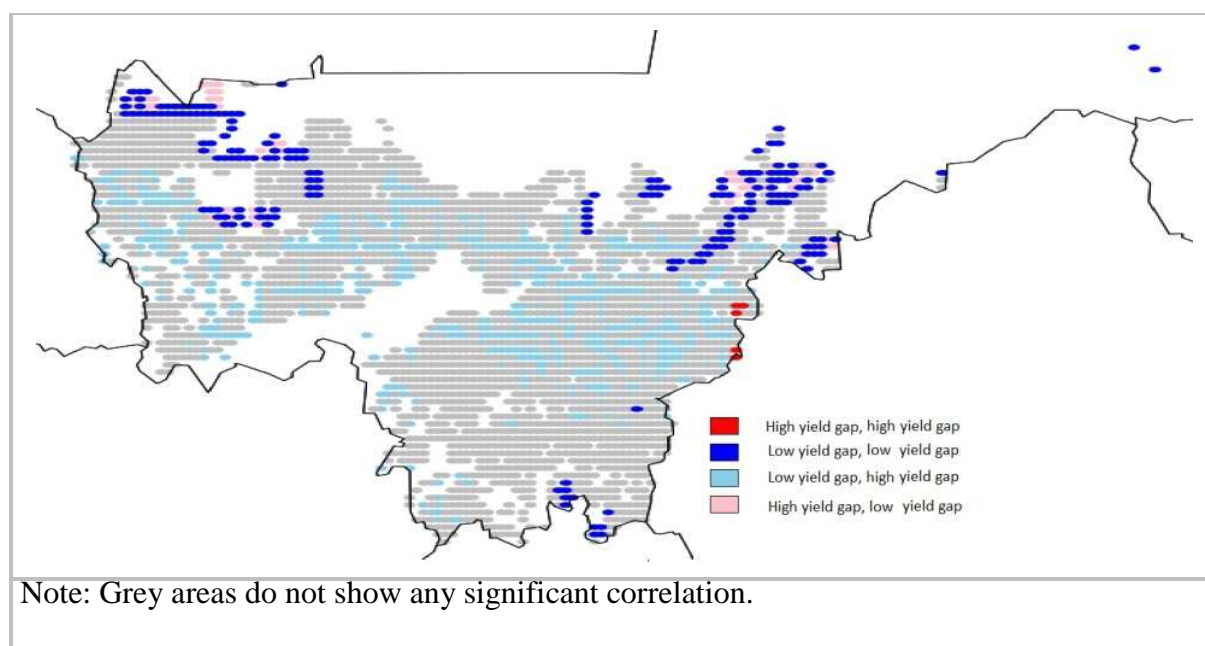


Figure 10 shows the hotspots and coldspots of large and small yield-gaps. Hotspots are areas where high yield gaps are clustered, i.e. there is a high correlation between an area (pixel) and its neighbouring areas (pixels). Large yield gap areas are characterised by a large difference between potential yield and actual yield. Coldspots are areas where low yield gaps are clustered. In areas with small yield gaps, the actual yields approach the potential yield. There are very few hotspots (where areas with large yield-gaps are adjacent to other areas with large yield-gaps) and there are a few coldspots (where areas with small yield-gaps are adjacent to other others with small yield-gaps). Most areas are mixed combinations of high and low yield gaps).

From the point of view of cost-effectiveness, it may be advisable to provide weather-index based insurance in dark blue area that are characterised by small yield-gaps. In these areas, the actual yields approach the potential yields. Crop failure will be determined by climatic factors, such as a lack of rainfall. In large yield gap areas, yields will vary because of lack of fertiliser use, for instance. In these regions it may be more difficult to disentangle the causes of crop failure and farmers may be disappointed when they do not receive indemnity even though their maize harvest failed.

Figure 11 Correlations between of yield gap and population density in southern Mali

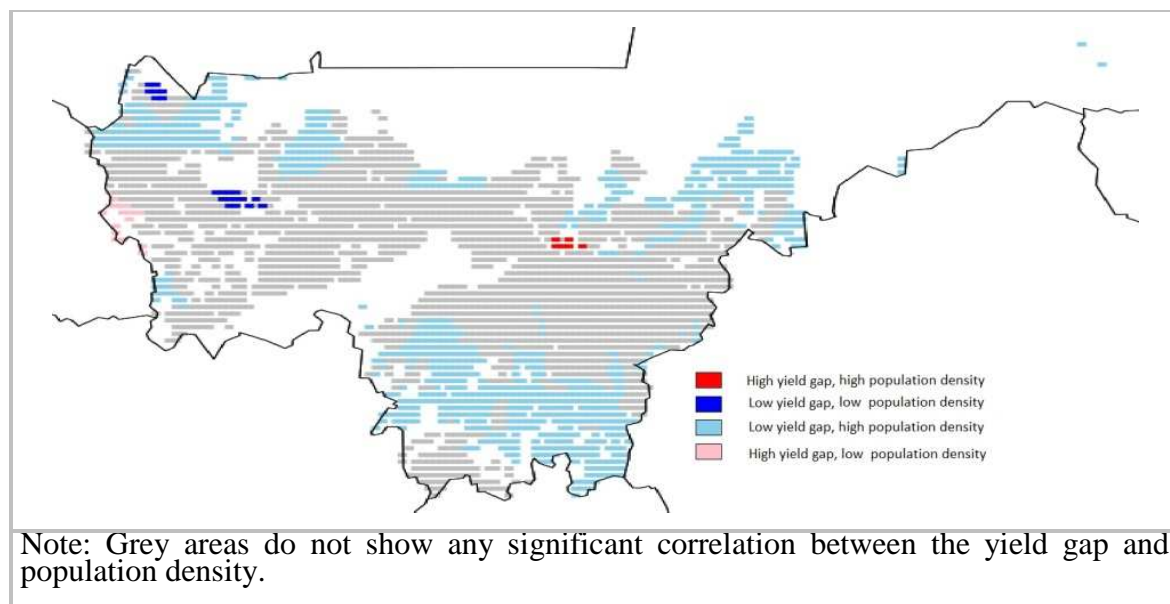
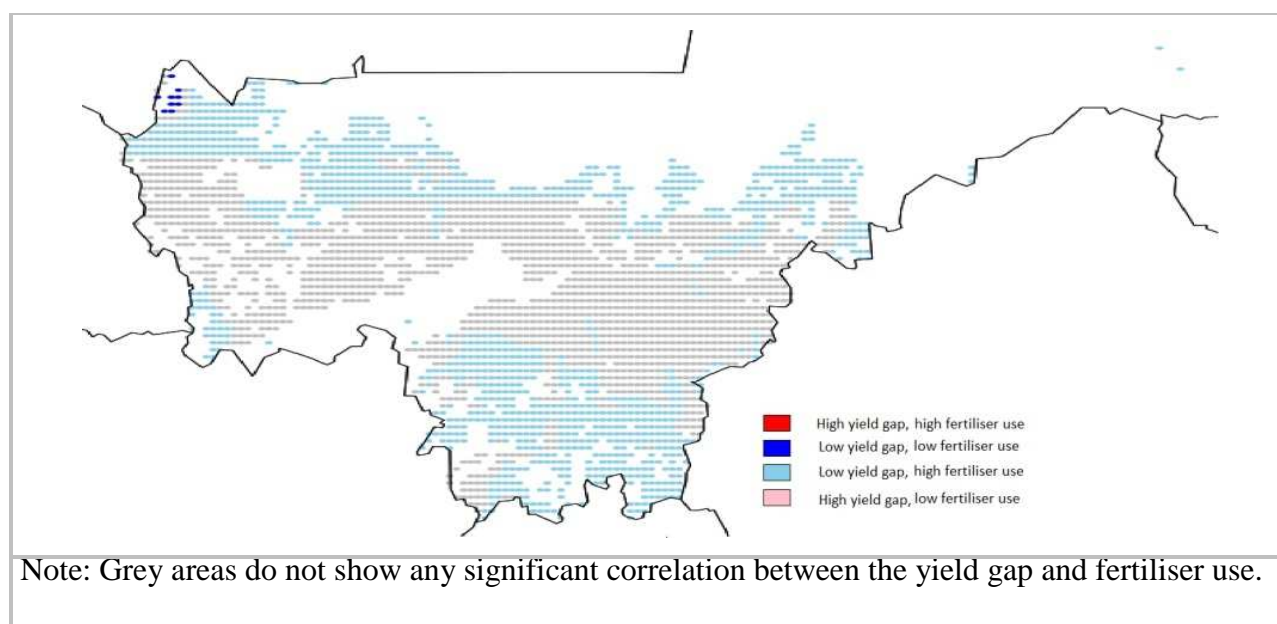


Figure 11 shows the correlations between yield gap and population density in Mali. If we compare the two light green areas from Figure 9 with Figure 11, we see that these mostly overlap with the grey areas that do not show any correlation. Population density does not play a role in explaining yield gaps in these regions.³

Figure 12 shows the correlation between yield gap and fertiliser use. About half the area is not statistically significant (grey), but the other half is characterised by a small yield-gap and high fertiliser use (pale blue). The light green area in Figure 9 only partly overlaps the pale blue in East Kayes (see for location Figure 9).

³ One noticeable area is the one characterised by low yield-gap and low population density and 90-100% variability (dark blue in Figure 11). Figure 4 shows that this is in a hilly area, and this might explain the uniqueness of this relatively small area. The potential yield is probably low, as is the population density. Variability may be high because areas with steep slopes need just the right amount of rain to produce a high yield (too much rain will lead to erosion).

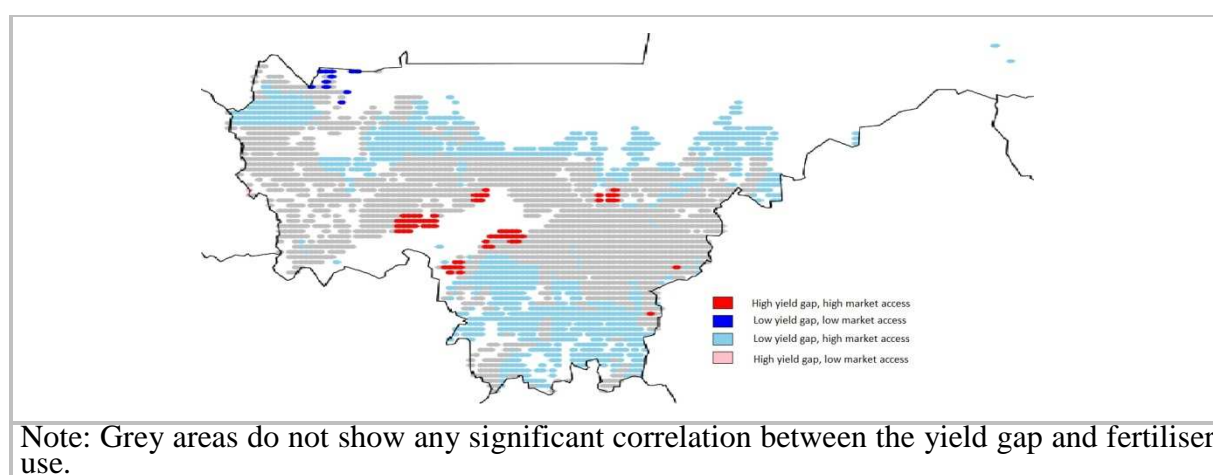
Figure 12. Correlations between yield gap and fertiliser use in Southern Mali



Not surprisingly, the pale blue in Figure 13 (small yield-gap and high market access) overlaps the pale blue in Figure 12 (small yield-gap and high fertiliser use): fertiliser use is partly determined by market access. There is again overlap with the light green area in Figure 9 (East Kayes).

Figure 13 has a few patches that are characterised by a seemingly contradictory combination of large yield-gap and high market access. They lie around Bamako, the capital city of Mali, which may explain both: good market access, but land use may not be geared towards growing maize.

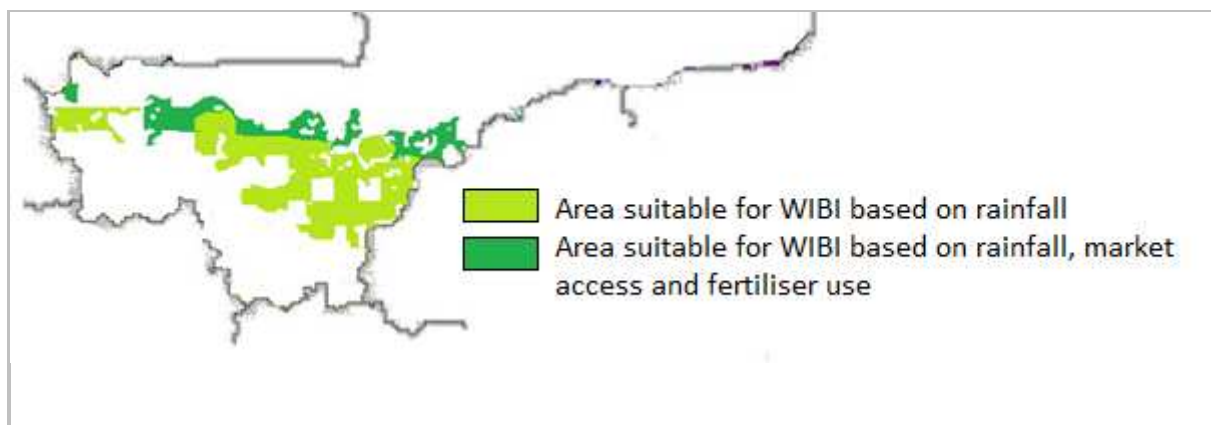
Figure 13. Correlations between yield gap and market access in Mali



When we distil the information from the previous pictures into Figure 14, we see that if we want to take into account market access and fertiliser use, a smaller area is suitable for weather index-based insurance (dark green) than the suitable range for yield variability (light

green) would suggest. We did not take into account population density here, as it was too scattered. Please note that the dark green areas were added by hand and are therefore approximations.

Figure 14. Areas suitable for weather index-based insurance (approximation)



5 Conclusions

This chapter has given a brief overview of how combining of spatial agro-ecological and socioeconomic data can be helpful in better identification of areas suitable for weather index-based insurance, taking into account various factors rather than only weather variability. Depending on the available data, different factors can be taken into account.

Combining different types of spatial data may be useful not only in targeting areas, but also in upscaling. If a pilot project has been running successfully in one area, it may be more easily upscaled to regions that have similar agro-ecological and socioeconomic characteristics. However, upscaling of insurance projects is not easy: contracts need to be carefully designed to fit the needs of the target population, taking into account specific risk factors that are important in that region. Transferring a pilot project to another area may mean recalculating the risks involved and redesigning the contracts. Mapping may help in finding similar areas for quick upscaling.

References

- Anselin, L., *An introduction to spatial autocorrelation analysis with GeoDa*. Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois, 2003.
- Anselin, L., 'Local indicators of spatial association - LISA.' In: *Geographical Analysis* 27 (1995) 2, pp. 93-115. doi:10.1111/j.1538-4632.1995.tb00338.x.
- Anselin, L., Sanjeev Sridharan and S. Gholston, 'Using exploratory spatial data analysis to leverage social indicator databases: The discovery of interesting patterns.' In: *Social Indicators Research* 82 (2006), pp. 287-309. doi: 10.1007/s11205-006-9034-x.
- Anselin, L., Ibnu Syabri and O. Smirnov, 'Visualising multivariate spatial correlation with dynamically linked window.' In: *New Tools for Spatial Data Analysis: Proceedings of a Workshop*, eds. Luc Anselin and S. Rey. Center for Spatially Integrated Social Science, Santa Barbara, 2002.
- De Bie, C.A.J.M., *Comparative performance analysis of agro-ecosystems*. ITC Dissertation, no. 75, 2000.
- Boserup, E., *The conditions of agricultural growth: The economics of agrarian change under population pressure*. Earthscan Publications, London, UK, 1965.
- Celebioglu, F. and S. Dall'erba, 'Spatial disparities across the regions of Turkey: An exploratory spatial data analysis.' In: *The Annals of Regional Science* 45 (2009) pp. 379-400.
- Chamberlin, J., J. Pender and B. Yu, *Development domains for Ethiopia: Capturing the geographical context of smallholder development options*. EPTD Discussion Paper. IFPRI, Washington DC, 2006.
- Conijn, J.G., E. Querner, M. Rau, H. Hengsdijk, T. Kuhlman, G.W. Meijerink, B. Rutgers and P.S. Bindraban, *Agricultural resource scarcity and distribution*. Research Report. PRI, part of Wageningen UR, the Netherlands, 2011.
- Conijn, J.G., H. Hengsdijk, B. Rutgers and R.E.E. Jongschaap, Mapping maize yield variability in Mali. PRI, part of Wageningen UR, the Netherlands, 2011.
- De Sherbinin, A., D. Balk, K. Yager, M. Jaiteh, F. Pozzi, C. Giri and A. Wannebo, *A CIESIN thematic guide to social science applications of remote sensing*. Palisades, Center for International Earth Science Information Network (CIESIN). Columbia University, NY, USA, 2002.
- Dorosh, P., Hyoung Gun Wang, Liangzhi You and E. Schmidt, 'Road connectivity, population and crop production in sub-Saharan Africa.' In: *Agricultural Economics* 43 (2012) January 1: pp. 89-103. doi: 10.1111/j.1574-0862.2011.00567.x.
- Ehui, S. and J. Pender, 'Resource degradation, low agricultural productivity and poverty in sub-Saharan Africa: Pathways out of the spiral.' In: *Agricultural Economics* 32 (January 1 2005): 225-242. doi:10.1111/0169-5150.2004.00026.x.
- Evans, L.T. and R.A. Fisher, 'Yield potential: its definition, measurement and Significance.' In: *Crop Science* 39 (1999) 6: pp. 1544-1551.
- FAO, 'How to feed the World in 2050.' In: *How to feed the world in 2050*. FAO, Rome, 2009. http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf
- FAO, *The state of food Insecurity in the world*. Rome, 2010. <http://www.fao.org/publications/sofi/en/>
- FAOSTAT, *FAOSTAT Database*. 2012a. <http://faostat.fao.org/>
- FAOSTAT, *World Hunger Map FAO*. 2012b. <http://faostat.fao.org/>
- Fischer, M.M. and A. Getis, *Handbook of applied spatial analysis: Software tools, methods and applications*. Springer, Berlin, Heidelberg, 2010.
- Fischer, R.A., D. Byerlee and G.O. Edmeades, 'Can technology deliver on the yield challenge to 2050.' In: *Expert meeting on how to feed the world*. Vol. 2050, 2009.
- Hodson, D.P. and J.W. White, 'Use of spatial analyses for global characterization of wheat-based production systems.' In: *Journal of Agricultural Science* 145 (2007): pp. 115-125.

- Jacquez, G.M., 'Spatial cluster analysis.' In: *The Handbook of Geographic Information Science*, by S. Fotheringham and J. Wilson, pp. 395-416. Blackwell Publishing, 2008.
- Kruseman, G., R. Ruben and G. Tesfay, 'Diversity and development domains in the Ethiopian highlands.' In: *Agricultural Systems* 88 (April 2006) 1: pp. 75-91. doi: 10.1016/j.agsy.2005.06.020.
- Licker, R., M. Johnston, J.A. Foley, C. Barford, C.J. Kucharik, C. Monfreda and N. Ramankutty, 'Mind the Gap: How do climate and agricultural management explain the 'yield Gap' of croplands around the World?' In: *Global Ecology and Biogeography* 19 (November 1 2010) 6: pp. 769-782. doi:10.1111/j.1466-8238.2010.00563.x.
- Lobell, D.B., K.G. Cassman and C.B. Field, 'Crop Yield Gaps: Their importance, magnitudes and causes.' In: *Annual Review of Environment and Resources* 34 (November 2009) 1: pp. 179-204. doi: 10.1146/annurev.environ.041008.093740.
- Meijerink, G. and K. Shutes, *Mapping socio-economic factors in Mali*. Nota. LEI, part of Wageningen UR, the Netherlands, 2011.
- Monfreda, C., N. Ramankutty and J.A. Foley, 'Farming the Planet: 2. Geographic distribution of crop areas, yields, physiological types and net primary production in the year 2000.' In: *Global Biogeochemical Cycles* 22 (March 1, 2008) 19 pp. doi: 200810.1029/2007GB002947.
- Müller, D. and M. Zeller, 'Agricultural intensification, population growth and forest cover change: Evidence from spatially explicit land use modelling in the Central Highlands of Vietnam.' In: *Land Use, Nature Conservation and the Stability of Rainforest Margins in Southeast Asia*. Springer-Verlag, Berlin, 2004.
- Neumann, K., P.H. Verburg, E. Stehfest and C. Müller, 'The yield gap of global grain production: A spatial analysis.' In: *Agricultural Systems* 103 (June 2010) 5: pp. 316-326. doi:10.1016/j.agsy.2010.02.004.
- Nin-Pratt, A., M. Johnson, E. Magalhaes, You Liangzhi, Xinshen Diao and J. Chamberlin, *Yield gaps and potential agricultural growth in West and Central Africa*. International Food Policy Research Institute, Washington, DC, 2011. <http://www.ifpri.org/publication/yield-gaps-and-potential-agricultural-growth-west-and-central-africa>.
- Omamo, S.W., Xinshen Diao, S. Wood, J. Chamberlin, Liangzhi You, S. Benin, U. Wood-Sichra and A. Tatwangire, *Strategic priorities for agricultural development in Eastern and Central Africa*. IFPRI Research Report. Washington DC, 2006.
- Páez, A., J. Gallo, R.N. Buliung and S. Dall'erba, *Progress in Spatial Analysis*. Advances in Spatial Science. Springer, Berlin Heidelberg, 2010.
- Pender, J., F. Place and S. Ehui, *Strategies for sustainable land management in the East African Highlands*. International Food Policy Research Institute, 2006.
- Pender, J.L., 'Rural population growth, agricultural change and natural resource management in developing countries: A review of hypotheses and some evidence from Honduras.' In: *EPTD Discussion Papers* 1999.
- Potter, P., N. Ramankutty, E. Bennett and S.D. Donner, 'Characterizing the spatial patterns of global fertilizer application and manure production.' In: *Earth Interactions* 14 (2) (January 2010): pp. 1-22. doi:10.1175/2009EI288.1.
- Ramankutty, N., A.T. Evan, C. Monfreda and J.A. Foley, 'Farming the Planet: 1. Geographic distribution of global agricultural lands in the year 2000.' In: *Global Biogeochemical Cycles* 22 (2008): 19 pp. doi:200810.1029/2007GB002952.
- Rau, M., T. Kuhlman and G. Meijerink, 'Why can't Africa produce more food? Mapping socio-economic constraints'. Poster presentation Tropentag, 5–7 October 2011, University of Bonn, Bonn, Germany.

- Richards, J.A. and Xiuping Jia, 2006. *Remote sensing digital image analysis: An introduction*. Springer. Berlin, Heidelberg, 2006.
- Siebert, S., P. Döll, J. Hoogeveen, J.M. Faures, K. Frenken and S. Feick, 'Development and validation of the global map of irrigation areas.' In: *Hydrology and Earth System Sciences Discussions* 2 (August 2005) pp. 1299-1327.
- United Nations, *World population prospects: The 2010 Revision*. 2011. <http://esa.un.org/unpd/wpp/index.htm>.
- Van Ittersum, M.K. and R. Rabbinge, 'Concepts in production ecology for analysis and quantification of agricultural input-output combinations.' In: *Field Crops Research* 52 (June 1997) 3: pp. 197-208. doi:10.1016/S0378-4290(97)00037-3.
- Verburg, P.H., E.C Ellis and A. Letourneau, 'A global assessment of market accessibility and market influence for global environmental change studies.' In: *Environmental Research Letters* 6 (2011) 3 (July 1). 034019. doi:10.1088/1748-9326/6/3/034019.
- Vollrath, D., 'A new look at agricultural productivity and economic growth.' Forthcoming in *Journal of Macroeconomics*. 2012.
- World Bank, *World Development Report 2008: agriculture for development*. World Bank, Washington DC, 2007.
- Zikhali, P., *Fast track land reform and agricultural productivity in Zimbabwe*. EfD discussion paper. Environment for Development, 2008.

Appendix 1: Crop growth simulation model LINPAC

The crop growth model used in the yield calculations originates from Spitters (1987) and Spitters and Schapendonk (1990) and is one of the so-called LINTUL (Linear INTERpolation of Utilization of Light) models (Bouman et al., 1996). Crop dry matter production in such models is calculated as the product of light interception and a light use efficiency (LUE, g dry matter per MJ intercepted radiation). Light interception depends on leaf area index (LAI, m² leaf per m² ground) and the accumulated dry matter production is distributed among above- and belowground and (non-)harvestable parts. Dry matter distribution is governed by the developmental stage of the crop (DVS, dimensionless) which is driven by temperature and can also be influenced by day length.

By combining the irrigation map of Siebert and colleagues (2005) and the cropland map of Ramankutty and colleagues (2008), the fraction of cropland equipped for irrigation was determined per grid cell at the 5 min arc resolution. With this fraction a weighted average yield was calculated per grid cell using the simulated potential and water-limited yields for maize. Data of cropland and irrigation refer to around 2000. For more information on how potential yield is calculated, we refer to Conijn and colleagues (2011a).

Two parameters that are regarded as key determinants of plant growth are used to distinguish 100 climate combinations: growing degree days (GDD) and a crop soil moisture index. GDD is a measure of the potential heat a plant can accumulate, and is normally calculated on a daily basis. It is defined as the number of temperature degrees above a certain threshold base temperature below which the plant is unable to grow. The base temperature varies among crop species. Final GDDs are computed by aggregating daily values over one year.

The crop soil moisture index is defined the annual average ratio of actual evapotranspiration to potential evapotranspiration. Evapotranspiration is the sum of evaporation (the movement to the air of water from sources such as the soil, canopy interception and water bodies) and plant transpiration (the movement to the air of water that vaporises from the leaves) from the Earth's land surface to the atmosphere.

GGD and the soil moisture index are each divided into 10 equal bins and combined to construct a 10 x 10 matrix that represent 100 climate zones. For each of these zones, maximum potential yield is defined as the 90 percentile of yield value for each climate zone. The cut-off point is introduced to avoid the use of outliers, which reflect potential erroneous data and might bias the results. The yield gap is eventually computed as the difference between the maximum yield potential by climate zone and the yield data from Monfreda and colleagues (2008).

Soil data

In this study the Digital Soil Map of the World (DSMW; FAO, 1995) is used for information on the spatial distribution of soil types in Mali. This gridded map has a resolution of 5x5 arcminutes in latitude/longitude coordinates and grid cells in Mali have an area of approximately 9x9 km (8100 ha). The legend of this global map contains 4,931 unique Soil Mapping Units (SMUs) and each grid cell is characterized by one SMU which can comprise a number of soil units (up to 8 different soil units per SMU). Mali comprises 89 different SMUs distributed over 15,269 grid cells that cover the whole country (1.25 million km² based on the sum of all grid cell areas). Most grid cells contain more than one soil unit and the areas of soil units are expressed as percentages of the grid cell. Each soil unit has unique soil properties, which may affect crop production. Selected soil properties comprise (1) texture class, (2) slope class, (3) soil depth, (4) available water content and (5) a soil induced reduction factor (the first four are derived from algorithms provided by the FAO). Available water content is defined here as the maximum amount of crop-available water that can be stored in the soil, i.e. the difference between field capacity (pF = 2.0) and permanent wilting point (pF = 4.2). The reduction factor limits crop growth and has been estimated as function of soil type in case of Acrisols (factor = 0.75), Solonetz (factor = 0.5) and if the phase description in the FAO soil database refers to saline/sodic conditions (factor = 0.25) to account for adverse soil chemical conditions.

Weather data

Weather data are obtained from the Climate Research Unit (CRU; <http://www.cru.uea.ac.uk/cru/data/>) and have a spatial resolution of 30x30 arcminutes (i.e. 54x54 km or circa 290,000 ha near the equator) and a temporal resolution of monthly values for each year in the period 1901 – 2100. The period 1901 – 2006 represents historical data and data from the period 2007 – 2100 are predicted values using a number of global circulation models and four emission scenarios (source: Tyndall Centre). In our analysis we have used historical data from 1976 – 2005 (30 years) comprising for each year in this period the monthly values of cloud cover, temperature, vapor pressure, rainfall and the number of rainy days (the so-called “CRU TS 3.0” database). Data on wind speed were not available per year and, therefore, we used average values obtained from the climatology of 1961 – 1990, also provided by CRU (“CRU CL 1.0”).

As the crop model runs with a daily time step, the monthly values of the weather database were linearly interpolated to obtain daily values for cloud cover, temperature, vapor pressure and wind speed. This procedure is not adequate for precipitation because it commonly consists of a series of discrete and random events. Using average daily precipitation values in the crop model may underestimate the effect of water stress on crop production. A random generator is therefore used to distribute the monthly total precipitation over the number of rainy days in a month, which results in a pattern of days without rain and days with a variable amount of rain in each month.

Land use data

We have used data from Monfreda et al. (2008) to derive spatially explicit information on the total harvested crop area and maize area in Mali at a resolution of 5x5 arcminutes. Their

global land use data are based on agricultural statistics from the period 1997 – 2003 in combination with remote sensing data.

Appendix 2: Spatial Regression Models

There are a number of approaches to the regression problem when one faces the problems associated with spatial autocorrelation. Detection of this problem is relatively simple. The Moran tests for a number of spatial specifications are possible with versions that are robust to spatial autocorrelation also reported. The data used are the yield gap data presented above. Estimation was performed using *spdep* in R.

Using an ordinary least squares (OLS) regression, residuals are retrieved and tested for spatial dependence. Thus the model is given by

$$YG_i = \alpha + \beta_1 Fert_i + \beta_2 Pop_i + \beta_3 Mark_i + \epsilon_i$$

where the usual properties of the OLS regressions are assumed to hold. Using the regression residuals, a test is performed to assess the type of spatial dependence in the model. The possibilities are a simple error dependence model and a spatially lagged dependent model, or a combination of the two.

The error dependence model assumes that the error term has a spatial structure, such that it can be written as

$$\epsilon_i = \lambda W\epsilon + u,$$

where λ is the spatial error parameter and W the spatial weighting. The variance of the residual vector is given by:

$$Var(\epsilon) = \sigma^2(I - \lambda W)^{-1}(I - \lambda W)^{-1}$$

From this the usual likelihood function can be derived; however, the logarithm of the determinant of the matrix $|I - \lambda W|$ is not easy to compute and may outweigh the sum of squares element of the likelihood function.

The alternative structure is that of an autoregressive lagged model of the form:

$$YG_i = \alpha + \beta_1 Fert_i + \beta_2 Pop_i + \beta_3 Mark_i + \rho WYG_i + \epsilon_i$$

where ρ is the spatial lag parameter. This specification is complicated by the fact that the ρ parameter feeds back into the system, making the interpretation more complex.

text above and below are not clear to me for different reasons. I understand that we may have two “problems”: (1) data may be spatially correlated (neighbours are more alike) and/or (2) explaining variables may be correlated (to each other).

Concerning (1): this may well be the situation because different resolutions have been used (5x5, 30x30, regions, countries?). Would that not ask for adjusting the resolution of 5x5 to a less detailed resolution? GIS tools can do that (i.e. finding the apparent/effective resolution)

Concerning (2): why has it not been investigated which model is the best in explaining the yield gap (X1, X2, X3, X1X2, X1X3, X2X3 and X1X2X3) Now only the last one (with 3 X

variables) is used but in case of correlated X variables, you can get “funny” results if two (highly) correlated X variables are used and leaving one out improves the explained variance.

In general it is not clear to me how above text links to below text, what the conclusion is (best model?), how this is used in the main text of section 4 (mapping socioecon. factors) and how this is related to yield & weather variability.

Basic linear regression

We use a basic linear regression to regress fertiliser use, population density and market access, with dependent variable yield gap. The results are presented in Table A1.1

Table A1.1	Basic linear regression for yield gap explained by fertiliser use, population density and market access		
	Estimate	S.E.	t-value
Intercept	5.97	9.19×10^{-3}	649.22
Fert	2.57×10^{-3}	3.33×10^{-4}	7.74
Pop	-1.18×10^{-4}	4.76×10^{-5}	-2.48
Mark	-1.36	4.44×10^{-2}	-30.68

Note that all coefficients are significantly different from 0 at 5%. The sign for fertiliser use is positive, which is counterintuitive. One would expect that more fertiliser use would lead to a smaller yield gap. The sign for population density is negative, which means that in highly populated areas, the yield gap tends to be smaller, which corroborates the induced innovation theory. Market access has a negative coefficient. With better market access, the yield gap tends to be smaller, which is to be expected.

Specification test

Using the forms of the models above, it is possible to test the restrictions that either $\alpha = 0$ or $\beta = 0$ is zero, or both are zero. Note that if both of the individual tests are significant, the robust forms should be used to select or guide model specification.

Table A1.2	Residuals of basic linear regression	
	Test Statistic	Df
LM error	280.875	1
LM lag	280.705	1
RLM error	444.8	1
RLM lag	274.7	1
SARMA	28.1150	2

It is clear from these tests that the OLS approach is poorly specified; the error model or the SARMA model might be the best approach to the modelling problem. Both the error and the lagged model are run and their results reported.

The regression models

Once the model is selected, the spatial correlation is estimated using an optimising algorithm with the regression following using a generalised least squares approach. One fundamental

problem with spatial regressions is the size of the various matrices, especially the weighting matrix W . A direct approach is often not feasible. Therefore, either a LU or Monte Carlo is used on the matrix $I - \lambda W$. A GMM model is possible in some cases.

Spatial models

The results of the spatial models are as follows.

The spatial error model assumes that the error carries the spatial dependence and thus estimates \square .

Table A1.3 Spatial error model for yield gap explained by fertiliser use, population density and market access			
	Estimate	S.E.	z-value
Intercept	5.66	$4.7156 \cdot 10^{-2}$	120.04
Fert	$5.76 \cdot 10^{-3}$	$8.01 \cdot 10^{-4}$	7.19
Pop	$7.68 \cdot 10^{-5}$	$2.61 \cdot 10^{-5}$	2.95
Mark	0.12	$4.86 \cdot 10^{-2}$	2.43
\square	0.92747		

The spatial lag model gives the following results.

Table A1.4 Spatial lag model for yield gap explained by fertiliser use, population density and market access			
	Estimate	S.E.	z-value
Intercept	$4.29 \cdot 10^{-1}$	$7.89 \cdot 10^{-3}$	54.34
Fert	$7.18 \cdot 10^{-5}$	$2.62 \cdot 10^{-5}$	2.74
Pop	$3.09 \cdot 10^{-5}$	$7.09 \cdot 10^{-6}$	4.36
Mark	$-9.98 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$	-4.21
\square	0.9265		

It should be noted that due to the impacts on the regression of the $\square \square$ the impacts need to be calculated and the coefficients are not interpreted in the same manner as those of OLS regressions. This is not the case in the spatial error model. Due to the matrix sizes, a Monte Carlo simulation was used to calculate these.

Table A1.5 Impact measures (lag, trace)			
	Direct	Indirect	Total
Fert	$9.74 \cdot 10^{-5}$	0,0008	0,0009
Pop	$4.20 \cdot 10^{-5}$	0,0003	0,0004
Mark	-0.14	-1.0846	-1.2200

The direct impact is the impact of changing the covariate on the dependent variable. The indirect effect accounts for the impact due to the neighbourhood effects. If the indirect impacts are large then there are significant spill-overs. In this case the indirect impacts are about 8 times larger than the direct; thus there are major spill-overs. There are possibly a number of reasons, for this but the most obvious one is that of externalities.

It is possible to estimate the lagged model using a form of two-stage least squares with spatially lagged X terms acting as instruments for the lagged dependent variable. This gives answers similar to those in the GMM estimation.

GMM models

The generalised method of moments can be used to estimate the regressions. The models are parallels of the error and lag maximum likelihood methods. The results are given below.

Table A1.6 GMM estimates of spatial error model			
	Estimate	S.E.	z-value
Intercept	5.70	3.69×10^{-2}	154.71
Fert	4.85×10^{-3}	7.57×10^{-4}	6.41
Pop	7.33×10^{-5}	2.64×10^{-5}	2.77
Mark	7.53×10^{-2}	4.87×10^{-2}	1.55
□	0.90515		

Table A1.7 GMM estimate of autoregressive model with spatially lagged dependent variable			
	Estimate	S.E.	z-value
Intercept	-4.57×10^{-1}	9.02×10^{-2}	-5.07
Fert	-3.68×10^{-4}	1.11×10^{-4}	-3.33
Pop	5.21×10^{-5}	1.68×10^{-5}	3.09
Mark	1.05×10^{-1}	2.58×10^{-2}	4.07
□	1.0750	1.51×10^{-2}	
□	-0.399		

As with the maximum likelihood estimation, the coefficients need careful interpretation. The calculations are given below.

Table A1.8 Impact measures (lag, trace)			
	Direct	Indirect	Total
Fert	-0.001	-0.037	-0.038
Pop	0.000	0.005	0.005
Mark	0.290	10.581	10.871

Appendix 3: Deriving the maize yield variability

Using a threshold of 25%, Figure 15 presents results for entire Mali, Figure 16 for the total harvested crop area in Mali and Figure 17 for the harvested maize area in Mali (see Conijn et al, 2011b). For example, the red coloured areas in these Figures indicate that in 80 to 90% of the years the yield is less than 25% of the average yield in these areas. Hence, this analysis provides insight in the spatial distribution of risk of (extreme) low yields (including crop failure) due to the (historical) weather conditions in combination with soil properties and crop (i.e. maize) characteristics.

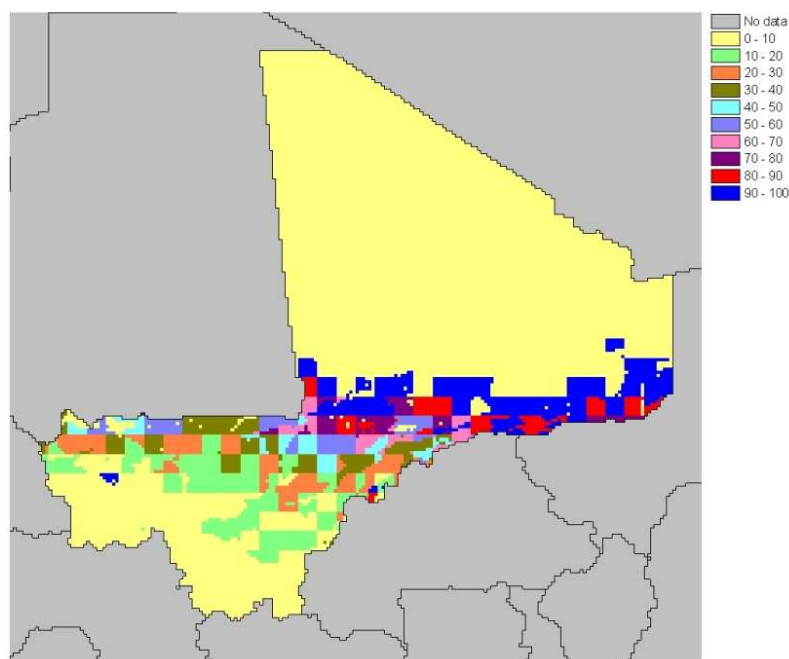


Figure 15. A map with percentage of years with low yield, here defined as < 25 % of the 30-years average rain fed yield per grid cell in entire Mali.

Large areas of low risks are in the North and South of Mali. In the North the risk is low because growing maize is not possible (not enough rainfall). In the South there are also areas with zero yields (due to soil conditions hampering crop growth), but most of the low risk in the South can be explained by the absence of high variability in weather conditions. Rains are more evenly distributed over the growing season and (extreme) low rain fed yields do not occur frequently (< 10%). Moving from the South-West towards the North-East the risk on low yields increases distinctively, illustrating the increase in variability of weather conditions combined with the soil characteristics that determine the soil water balance and the sensitivity of maize to drought. The risks of the area that lies just below the zero yield region in the North of Mali are very high (> 90%: dark blue colour). These high risks are associated with a low frequency of years with sufficient rainfall to allow a maize cropping season, because in only a few years (< 3) rainfall is adequate to grow maize. It is expected that these areas are not used for cropping or maize growing unless irrigation is possible.

Figure 16 and Figure 17 are subsets of Figure 15 and they show the same risks for only the total harvested crop area (Figure 16) and the harvested maize area (Figure 17) in Mali. By focusing on grid cells with harvested maize areas (Figure 17), the results of the crop growth model (supplied with maize growth parameters) are linked to those areas where maize is actually cultivated in Mali and risks of maize cultivation can be analysed more precisely.

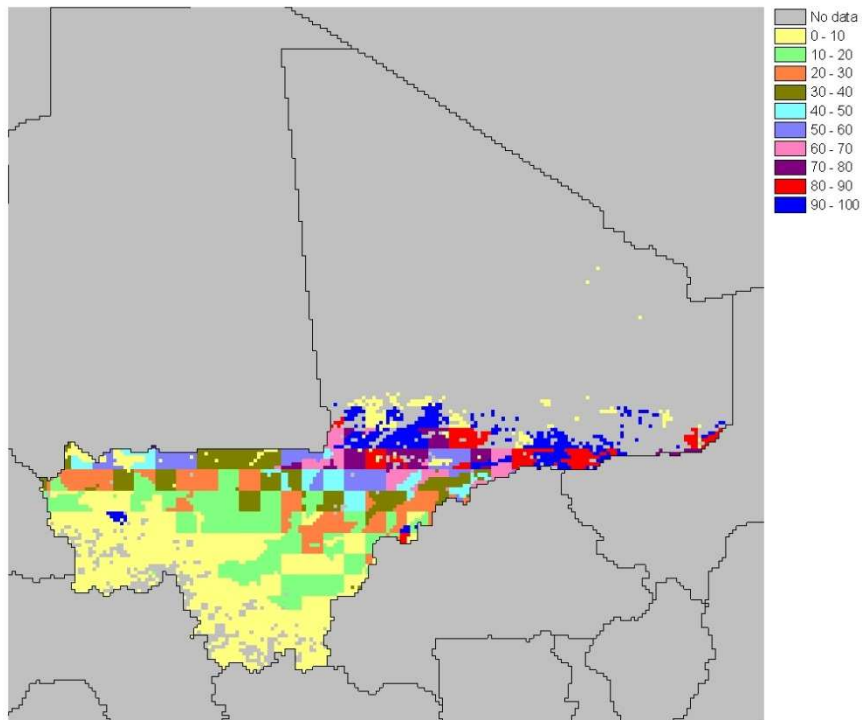


Figure 16. A map with percentage of years with low yield, here defined as $< 25\%$ of the 30-years average rain fed yield per grid cell used for growing arable crops in Mali around the year 2000 (= subset of figure 15). Grey areas in Mali refer to areas without arable crops (Monfreda et al., 2008).

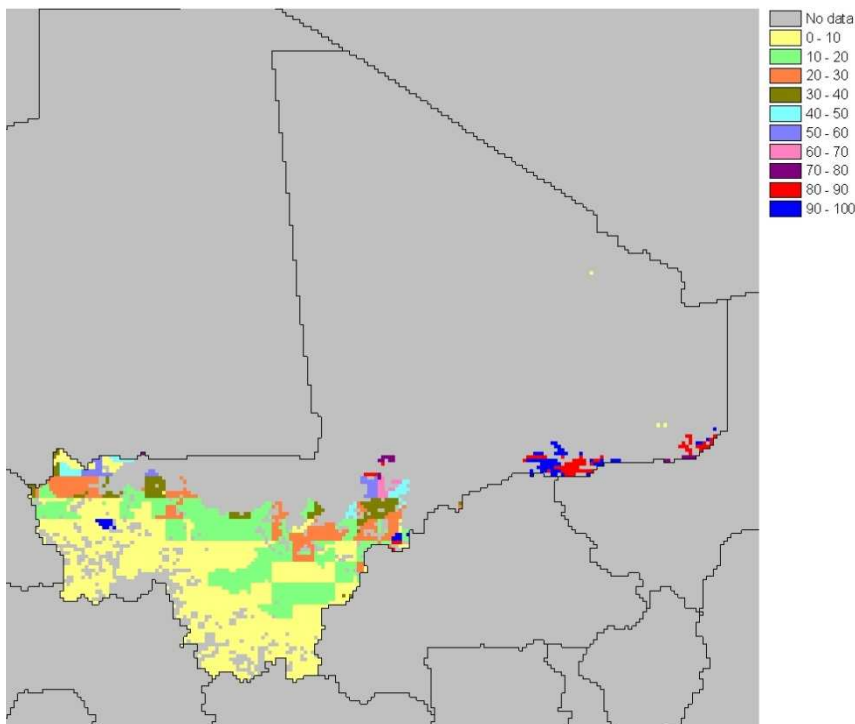


Figure 17. A map with percentage of years with low yield, here defined as $< 25\%$ of the 30-years average rain fed yield per grid cell used for growing maize in Mali around the year 2000 (= subset of figure 16). Grey areas in Mali refer to areas without maize cultivation (Monfreda et al., 2008).

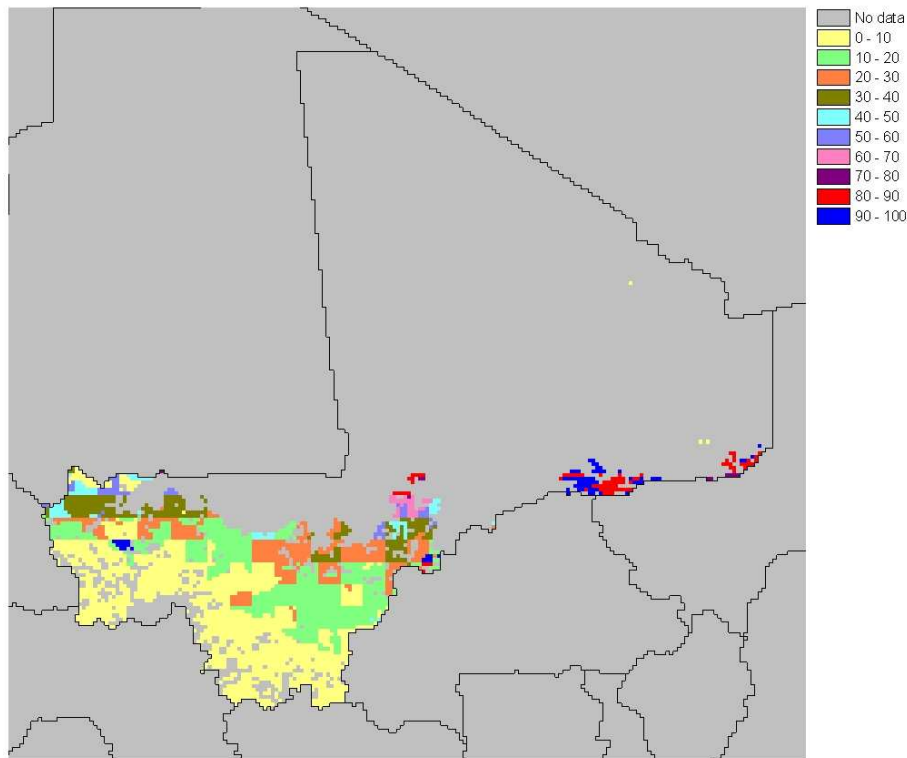


Figure 18. A map with percentage of years with low yield, here defined as < 50% of the 30-years average rain fed yield per grid cell used for growing maize in Mali around the year 2000 (compare with Figure 17).

Spatial risk distribution as illustrated in Figures 15-17 depends on the selected yield threshold (i.e. < 25% of average grid cell yield). Figure 18 gives a similar map as Figure 7 using a threshold of < 50% of the average yield per grid cell.